A Machine Learning Model for Pulmonary Embolism Prediction Using a Large Multi-Center Database

L. TEWALA[1], H. KIM[2] and R. D. STEVENS[2]

[1] Department of Biophysics, Krieger School of Arts and Sciences, [2] Department of Anesthesiology and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, USA

INTRODUCTION

- Pulmonary embolism (PE) is a frequent and life-threatening complication in critically ill patients.
- While PE risk scores have been proposed, there is an unmet need for accurate methods that predict the likelihood for PE.

OBJECTIVES

- Build a data-driven computational model to predict PE in patients admitted to the ICU
- Analyze this model to understand the most important prediction features

METHODS

- Utilized a multisite clinical database of 208 US institutions, the eICU collaborative research database
- Identified 2100 patients, half of which were admitted to the ICU and then developed PE
- For the PE patients, used data from ICU admission till the patient was diagnosed with PE
- For the non-PE patients, used data from their first 6 hours in the ICU
- Trained three different machine learning (ML) algorithms: random forest (RF), gradient boosting (XGBoost), and generalized linear models (GLM)
- Started with 286 demographic and laboratory features
- Used forward selection to obtain the most efficient feature space of 69 features

RESULTS

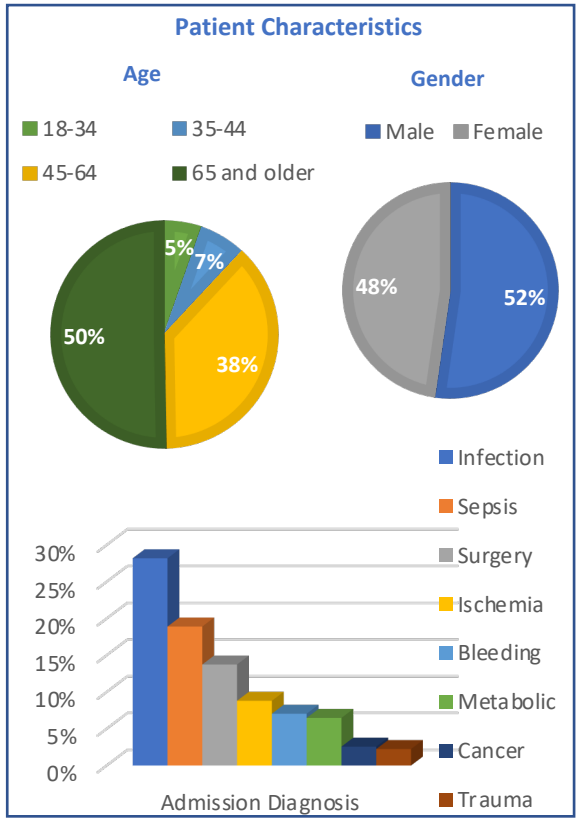
- The best performing model was the random forest model with an AUROC of 0.91 ± 0.01.
- Highly ranked features **directly** proportional with PE diagnosis include final and mean partial thromboplastin time (PTT), final prothrombin time (PT), mean basophil count, and final white blood cell count (WBCx1000).
- Highly ranked features **inversely** proportional with PE diagnosis include mean anion gap, initial PTT, initial blood urea nitrogen (BUN), and initial international normalized ratio PT (PT/INR).
- This study also plans to explore time-dependent physiological variables and PE patients with recorded CT scans.

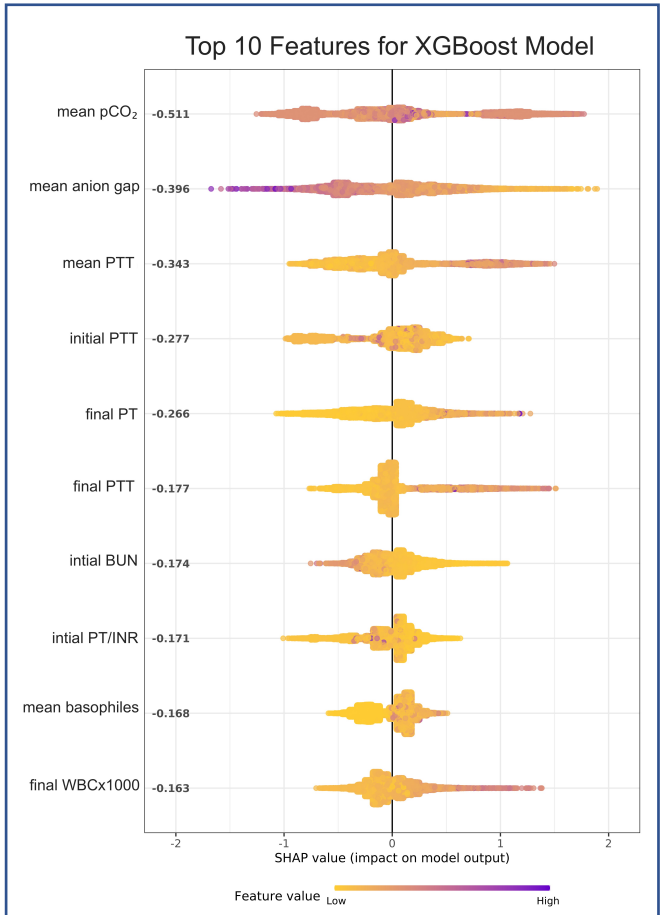
2906 PE Patients in the eICU database → **1299** With no PE/DVT admission diagnosis → **1050** With sufficient feature data

Model	AUROC	Sensitivity	Specificity
RF	**0.91 ± 0.01**	**0.86 ± 0.03**	**0.84 ± 0.03**
XGBoost	0.90 ± 0.01	0.84 ± 0.03	0.82 ± 0.03
GLM	0.86 ± 0.02	0.82 ± 0.02	0.75 ± 0.03

CONCLUSIONS

- Results demonstrate the potential value of ML models using clinical variables to help predict the onset of PE in the ICU.
- Following external validation and testing against the established risk scores, this study aims to produce a model that could accurately predict PE in real-time in the ICU.





Negative and positive SHAP values indicate reduced and increased likelihood of PE respectively.